Performance metrics (classification)

Jacobo Robledo

2024-10-24

Here are definitions of the performance metrics I presented in Classification\_Notebook.html when comparing machine learning (ML) models:

### 1. Accuracy

* **Definition**: The proportion of correctly classified instances (true positives and true negatives) out of the total instances.
* **Formula**:
* **Purpose**: Measures the overall correctness of the model.

### 2. AUC (Area Under the Curve)

* **Definition**: The Area Under the Receiver Operating Characteristic (ROC) curve. It measures the ability of a model to distinguish between positive and negative classes.
* **Purpose**: AUC represents how well the model discriminates between the positive and negative class. A higher AUC indicates better performance.

### 3. Kappa (Cohen’s Kappa)

* **Definition**: A statistical measure of inter-rater agreement or classification accuracy, adjusted for chance.
* **Formula**:
* **Purpose**: Takes into account both the accuracy and the possibility that agreement could happen by chance.

### 4. LogLoss (Logarithmic Loss)

* **Definition**: Measures the performance of a classification model where the output is a probability value between 0 and 1. It penalizes false classifications with a focus on how confident the model was in making the wrong prediction.
* **Formula**:
* **Purpose**: Lower log loss means better performance. It evaluates the uncertainty of predictions.

### 5. Mean Balanced Accuracy

* **Definition**: The average of recall obtained on each class, useful for imbalanced datasets.
* **Formula**:
* **Purpose**: Corrects the bias introduced by imbalanced datasets.

### 6. Mean Detection Rate

* **Definition**: The proportion of positive cases correctly detected by the model out of the total instances. It’s typically equivalent to sensitivity or recall.
* **Formula**:

### 7. Mean F1 (F1 Score)

* **Definition**: The harmonic mean of precision and recall. It gives a balance between precision and recall.
* **Formula**:
* **Purpose**: Best used when you need a balance between precision and recall.

### 8. Mean Negative Predictive Value (Mean NPV)

* **Definition**: The proportion of negative cases correctly identified by the model out of all predicted negative cases.
* **Formula**:
* **Purpose**: Useful for assessing the ability of a model to correctly rule out negative cases.

### 9. Mean Positive Predictive Value (Mean PPV)

* **Definition**: The proportion of positive cases correctly identified by the model out of all predicted positive cases (i.e., **Precision**).
* **Formula**:

### 10. Mean Precision

* **Definition**: The average of precision scores across different classes. Precision is the proportion of true positives among all predicted positives.
* **Purpose**: Indicates how reliable the model is when it predicts a positive class.

### 11. Mean Recall

* **Definition**: The average recall (or sensitivity) across classes, representing the ability of the model to capture all true positives.
* **Formula**:
* **Purpose**: Reflects the model’s ability to find all relevant instances.

### 12. Mean Sensitivity

* **Definition**: Another term for **Recall**, which measures the proportion of actual positives that are correctly identified by the model.
* **Purpose**: Sensitivity is important in detecting positive cases.

### 13. Mean Specificity

* **Definition**: The proportion of actual negatives that are correctly identified by the model.
* **Formula**:
* **Purpose**: Measures the model’s ability to detect true negatives.

### 14. prAUC (Precision-Recall AUC)

* **Definition**: The area under the Precision-Recall curve. It focuses more on the performance of a model on the positive class and how well it balances precision and recall.
* **Purpose**: Useful for imbalanced datasets where focusing on the positive class is important, as it emphasizes how well the model captures true positives while avoiding false positives.